The Spanish Journal of Psychology
2011, Vol. 14 No. 1, 207-217
doi:10.5209/rev_SJOP.2011.v14.n1.18

ISSN 1138-7416

Are We more Consistent when Talking about Ourselves than when Behaving? Consistency Differences through a Questionnaire and an Objective Task

Víctor J. Rubio, José Manuel Hernández, Javier Revuelta, and José Santacreu

Universidad Autónoma de Madrid (Spain)

The present paper aimed to examine questionnaire response patterns and objective task-based test behavioral patterns in order to analyze the differences people show in consistency. It is hypothesized that people tend to be more consistent when talking about themselves (when describing themselves through verbal statements) that when they solve a task (when behaving). Consistency is computed using the π^* statistic (Hernandez, Rubio, Revuelta, & Santacreu, 2006). According to this procedure, consistency is defined as the value and the dimensionality of the latent trait of an individual (θ) remaining invariant through out the test of. Participants who are consistent must show a constant θ and follow a given response pattern during the entire course of the test. A sample of 3,972 participants was used. Results reveal that 68% of participants showed a consistent response pattern when completing the questionnaire. When tackling the task-based test, the percentage was 66%. 45% of individuals showed a consistent pattern in both tests. Implications for personality and individual differences assessment are discussed.
Keywords: personality, consistency, prediction of behaviour, self-reports, objective task-based personality tests.

El presente artículo pretende examinar las diferencias que muestran las personas en cuanto a su consistencia entre los patrones de respuesta a un cuestionario y los patrones de respuesta ante un test objetivo basado en tareas. Se hipotetiza que las personas tienden a ser más consistentes cuando hablan sobre ellos mismos (es decir, cuando se describen en base a declaraciones verbales, como es el caso de los cuestionarios) que cuando resuelven una tarea (cuando se comportan). La consistencia se calcula utilizando el estadístico π^* (Hernandez, Rubio, Revuelta, & Santacreu, 2006). De acuerdo con este procedimiento, la consistencia se define como la invarianza del valor y la dimensionalidad del rasgo latente de un individuo (θ) a lo largo de un test. Los participantes que son consistentes mostrarán una θ constante y seguirán un patrón de respuesta dado a lo largo del curso completo del test. Para este estudio se utilizó una muestra de 3972 personas. Los resultados muestran que el 68% de los participantes mostraron un patrón de respuesta consistente cuando cumplimentaron el cuestionario. Sin embargo, cuando se enfrentaron al test objetivo basado en tareas, el porcentaje fue del 66%. El 45% de los participantes mostraron un patrón consistente en los dos tests. El artículo analiza las implicaciones para la evaluación de la personalidad y las diferencias individuales.
Palabras clave: personalidad, consistencia, predicción de la conducta, autoinformes, test objetivos basados en tareas.

This research has been funded by the Spanish Ministerio de Ciencia y Tecnología (MCyT, BSO2003-02162) and by the Spanish Ministerio de Ciencia e Innovación (PSI2008-02916).

Authors would like to thank Prof. M. Angeles Quiroga who read a previous version of the manuscript and made interesting suggestions for its improvement.

Correspondence concerning this article should be addressed to José Santacreu. Departamento de Psicología Biológica y de la Salud. Universidad Autónoma de Madrid. Campus Cantoblanco. 28049 Madrid. (Spain). Phone : +34-914975179. Fax: +34-914975215. E-mail: jose.santacreu@uam.es

Individuals' consistency refers to a relatively enduring behavioural disposition. Individuals are consistent if they behave in a similar way in situations pertaining to the same category (i.e. situations which elicit anxiety and/or fear, situations which involve risk, co-operation situations, etc.). Temporal and situational consistency in behavior allow human beings to see themselves as being unique and different (Pervin, 1996).

Nevertheless, in spite of the assumption that every individual is consistent in situations related to the same personality dimension, an assumption which lies at the core of most definitions of personality traits (West & Graziano, 1989), there are several concerns surrounding this. For instance, can it be accepted that personality consistency is stable along the life span of individuals, at least from adulthood, or is there a peak of trait consistency before and after when human beings are more prone to exhibit changes? Is it possible to produce personality changes by means of psychological treatment? Is every personality dimension as consistent and stable as the others? (Conley, 1984; Hellervik, Hazucha, & Schneider, 1992; Linehan & Kehrer, 1993; McCrae & Costa, 1994; Roberts & DelVecchio, 2000). In sum, can it be stated that every individual is equally consistent, is always consistent, and is consistent across all of the dimensions?

Controversy regarding consistency has been on the front page of many psychological journals since Mischel's contentions. As is well known, Mischel (1968), in light of the review of different research, pointed out that results do not support the hypothesis that individuals are consistent. To the contrary, he stated that individuals' behavior is situationally determined. Arguments against Mischel's suggestion were posed from different perspectives and authors (see Bem & Allen, 1974; Epstein, 1985; Funder, 1983; Ozer, 1986). Mischel himself would modify his approach (Mischel, 2004; Mischel & Shoda, 1998; Shoda & Mischel, 2000).

Probably the most fruitful result of the so-called "consistency debate" has been the extension of the notion of consistency. For instance, as Roberts and DelVecchio (2000) wrote, there are various forms of consistency: intraindividual differences in consistency (absolute consistency), ipsative consistency, mean-level consistency, and rank-order consistency; some of them focused on individuals and others focused on groups. Nevertheless, there is still a gap between theory and common psychological practice when a person is assessed using some of the instruments available, or prior to this when a personality assessment instrument is developed. In such cases, it is assumed that the individual is consistent and stable. From a (classical) psychometric point of view, a good instrument for assessing his/her personality traits is one made up of elements that can detect the situational consistency of individuals irrespective of the magnitude of the trait variable in them. Consequently, the process of constructing an assessment instrument that aims to measure personality traits demands the elimination of those items that do not contribute to increasing the internal consistency of the test, as it is assumed that those items are not measuring the dimension or trait being explored. That is, it is understood that an individual is consistent across the different situations represented by the test items. However, in spite of the fact that the sources of variation in the scores of a test are two-fold (a) the items of the test and b) the individuals tested, methodological and analytical procedures have not been implemented to allot separate values to individual consistency and to the consistency of the elements that make up the test. On the same line, the estimation of the proportion of variance of the empirical scores due to the true score variance is made by the degree of temporal consistency of subjects' scores through successive replications of the test. When using a test-retest replication design for establishing the reliability of test scores and significant differences appear between the two administrations (which have supposedly controlled the time between tests, the test conditions, the examiner, the subjects' developmental issues, etc.), such differences are usually attributed to the instrument. Given the fact that no changes in the elements have been produced either, there is no reason to rule out the possibility of an individuals' lack of temporal consistency.

Consistency links directly with the possibility to make predictions about future behaviour when circumstances remain constant. However, the problem regarding the capacity to make accurate predictions about individuals' future behaviour is not solved. Even though there is a significant amount of data which demonstrates a correlation between personality dimensions (usually, all or some of the big five) and constructs such as happiness, distress or marital satisfaction, predictive power over specific behaviours is dramatically reduced. It decreases, in turn, the usefulness of personality variables as predictors in specific contexts (Paunonen & Ashton, 2001a, 2001b; Paunonen, Haddock, Forsterling, & Keinonen, 2003; see Farsides & Woodfield's, 2003 review about academic achievement and personality or Barrick & Mount's, 1991; Salgado's, 1997; Schmidt & Hunter's, 1998; Tett, Jackson, Rothstein, & Reddon's, 1994 meta-analyses regarding occupational achievement and predictive capacity of different personnel selection instruments).

The relationships between what people say and how they really behave

One possible source of difficulty in making predictions about complex future behaviours has been related to the assessment instruments used for personality measurement. Several authors have verified that people's verbal statements do not always match their actions (Cattell, 1965; Lang, 1971; Skinner & Howarth, 1973; Wagerman & Funder, 2007). Pawlick (1985) focused on self-report

https://doi.org/10.5209/rev_SJOP.2011.v14.n1.18 Published online by Cambridge University Press

biases such as errors of retrospection, errors of quantitative or qualitative levelling and sharpening, or errors of self-attribution. However, other questions not directly related to self-report biases should be taken into consideration. Firstly, the role of language when describing oneself and others. Semin and Krahe (1988) studied the mediating role of language when lay persons establish relationships among behaviours when describing the personality of an individual. They found that, in the absence of a pragmatic context, behavioral descriptions can only be understood with reference to an abstract, semantic context consensually shared by the language speakers. As Semin and Krahe pointed out, when individuals have to make judgments based on verbal descriptions presented in an abstract and decontextualizated way, those judgments are made with reference to linguistic conventions instead of intuitive notions about personality. This phenomenon would explain why self-report instruments assessing the same dimension usually show high correlations, in contrast to what happens when using behavioural measures (Zuckerman, 1979; see Bromley & Curley, 1992, for an analysis of this phenomenon applied to risk tendency assessment).

Secondly, several authors have suggested a dual-system model for describing the differences between propositional vs. automatic processes (Smith & DeCoster, 2000; Strack & Deutsch, 2004). Even though such models have not been developed for assessment purposes, they have recently been used to take into account the lack of convergent validity of indirect vs. direct measures (for instance, see Hofmann & Schmitt, 2008). From this point of view, it is suggested that procedures based on direct assessment via self-report of personality dimensions tap into different modules of the information processing system to those based on implicit or indirect assessment (overt questioning about usual behavioural tendencies or attitudes) (for instance, see Rudolph, Schröeder-Abe, Schütz, Gregg, & Sedikides, 2008).

In the same way, Santacreu, Rubio, and Hernández (2006) have pointed out that the inconsistencies between what people say about themselves and what they do is related to the fact that the two ways of synthesizing the experience are different. If a person says he/she is tidy but on his/her desk there are piles of books and documents, this does not necessarily mean that the person is lying. If an individual is asked, he/she would certainly answer that over the last few days he/she has been very busy and papers were mounting up. Or the person would answer that, believe it or not, everything is in its place and he/she would be able to find any document. Whatever the individual says, it demonstrates that the way a person qualifies him/herself is different to the way that person behaves and that should have a direct effect on the consistency shown by one or other kind of manifestation. When making verbal statements about him/herself, a person tries to be coherent (Cervone & Shoda, 1999; Greenwald, 1980; Ross, 1989). That is, people try to avoid contradictions in their responses and try to preserve a "sense of consistency" about themselves, organizing their experiences into a meaningful life story (McLean, Pasupathi, & Pals, 2007). It should be noted that these verbal descriptions about oneself are usually produced, at least at the beginning, at the request of others (when someone is asked) or have been assimilated from the statements that others make about an individual. This sort of consistency does not need to be maintained when "behaving".

Estimating individual consistency on test scores

Mischel's (1968) criticism about the lack of empirical evidence of the consistency supposition brought about the distinction, initially pointed out by Bem and Allen (1974), between trait level and what was later called traitedness. Traitedness refers to the relevance of a personality trait to an individual's personality (see, for instance, Cucina & Vasilopoulos, 2005). In the end, traitedness moderates the relationships between personality test scores and behaviour. Estimating the individuals' traitedness, however, becomes a methodological problem. The first approach consisted of asking individuals about how relevant the trait is for them. This is the case of the Bem and Allen's (1974) Cross Situation Behavior Survey (CSBS), for instance. However, this approach has been questioned due to reasons such as social desirability bias (Rushton, Jackson, & Paunonen, 1981), ego-defensive attributional bias (Weiner, 1990), or the potential confound between ratings of consistency/reliability of behaviours relevant to a trait, and the trait level, which show a curvilinear relationship (extreme scores show less variability) according to Paunonen and Jackson (1985).

There have been other approaches of measuring traitedness, although they are based on the self-report measure, such as that offered by Chaplin (1991) or Biesanz and West (2000), who proposed the *construct similarity*; the correlations between individual's estimations about a set of items which measure a trait and the mean of estimations of the rest of the sample about the same items. The construct similarity tries to differentiate between individuals whose trait interpretation is similar to others' who have an idiosyncratic interpretation.

Bem and Allen (1974) also proposed another meaure of traitedness: the *ipsatized variance index* (IVI). IVI is based on the ratio of the variance of item responses within a scale of the variance of item responses for the entire test. It means that the IVI should be obtained using multi-trait questionnaires. Further to these, some other measures based on Bem and Allen's IVI have been suggested (the inter-item variance –Baumeister & Tice, 1988; the inter-item variability – Biesanz & West, 2000). These procedures have been criticized due the lack of theoretical bases (Paunonen & Jackson, 1985). Baumeister and Tice (1988) also pointed out that the quotient in which IVI is based is not necessary

https://doi.org/10.5209/rev_SJOP.2011.v14.n1.18 Published online by Cambridge University Press

for describing the consistency of the responses. On the same line but using a more sophisticated procedure is the Lanning's (1988) *scalability* method which estimates traitedness computing the absolute value of residuals obtained, comparing observed scores with expected scores regarding the variability of individuals and items.

However, as Cucina and Vasilopoulos, (2005) have pointed out, many of the traitedness measures did not show adequate divergent validity from trait level and little evidence of convergent validity among different measures have been demonstrated.

Hernández, Rubio, Revuelta, and Santacreu (2006) developed a new statistic to estimate individual`s consistency. The π* statistic is an adaptation of the π* statistic proposed by Rudas, Clogg, and Lindsay (1994) for the analysis of contingency tables. It interprets consistency as the invariance throughout the test of the value and the dimensionality of the latent trait of an individual (θ). Participants who are consistent must show a constant θ and follow a given response model during the entire course of the test. The probability of sampling a consistent or non-consistent individual at random from each group is $1 - \pi^*$ and π*, respectively. $X_h = (X_{h1}, ..., X_{hn})$ is the participants response pattern to a test of I items. H (h = 1, ..., H) is the number of different response patterns. The probability of each response is given by an Item Response Theory (IRT) model (Birnbaum, 1968). The conditional probability that a subject is consistent is

$$f(\text{consistent}|X_h) = [(1 - \pi^*) P(X_h)/f(X_h)] n_h/n \quad [1]$$

where

- X_h: response pattern
- $1 - \pi^*$: probability of randomly sampling a consistent participant
- $P(X_h)$: probability of responses that follow the model
- $f(X_h)$: marginal probability of the response pattern of the participant
- n_h: observed response frequency of the pattern h
- n: number of participants

Hernandez et al. (2006) demonstrated the advantages of π* compared to l_z of Drasgow, Levine, and Williams, (1985) and to α of Ferrando (2004). π* is independent to the model. Thus, it can be computed with any model that specifies how consistently participants might respond, and provides an estimation of how many participants are consistent in a given population. Instead, l_z statistic must be verified participant by participant, a procedure that increases the probability of a Type I error. In contrast to Ferrando's α, π* does not assume that all participants are inconsistent, nor does it assume anything about how inconsistently individuals act.

According to the π* goodness of fit index, Hernández et al. (2006) showed an interesting 36%, approximately, of inconsistent individuals, higher that what would have been expected from an assumption of every individual being consistent. Two objective task-based personality tests were used to assess thoroughness and risk propensity. Objective task-based personality tests used in that paper are based on procedures designed by Cattell (see Cattell, 1965, Cattel & Kline, 1977; Cattell & Warburton, 1967) for assessing the so called T-data, instead of Q-data. Q-data are the responses to questionnaires in which individuals report about themselves. T-data are those obtained by tasks in which individuals are not aware of the specific dimension which is measured. These procedures consist of standard situations in which response options and/or the way an individual performs the task are registered. Besides keeping the dimension assessed masked, objective task-based personality tests do not require self-reports of the individual, nor is interpretation by the assessor needed (unlike projective tests). Furthermore, the tasks used do not require a high level of skill, so that the differences found are due to the personality dimension and not to the individual's abilities. Moreover, following the proposal of Caspi and Moffitt (1993), who suggested that personality differences are more likely to appear in transitions to new situations characterized by their unpredictability in which there is pressure to behave but a lack of information about how one might behave in an adaptive way, no feedback about performance is provided so that individuals cannot learn the "right response" (Santacreu et al., 2006) and the situation remains novel and ambiguous in Caspi and Moffitt's (1993) sense. Therefore, people behave as their dispositional tendencies would impose.

With the aim of analyzing individual consistency people show in a self-report based instrument as well as in an objective task-based instrument, the present papers attempts 1) To test whether individuals show consistency in their answers to a personality questionnaire regardless of internal consistency of items and global test scores using the π* statistic, as well as to replicate the previous results obtained (Hernandez et al., 2006) in an objective task-based test. 2) To compare individual consistency shown in Q-data vs. consistency shown in T-data by the same subjects using a questionnaire for assessing Emotional Stability and a task-based test for assessing Thoroughness.

Method

Participants

3,792 Spanish university graduates who applied for a selection process for a highly qualified position related to air transportation. All of them were native Spanish speakers as well as highly fluent in English. 2,111 (55.6%) were males and 1,681 (44.4%) were females. Mean age was 29.33 (*Sd* = 3.88).

https://doi.org/10.5209/rev_SJOP.2011.v14.n1.18 Published online by Cambridge University Press

Materials

*Emotional Adjustment Bank,*_EAB (Aguado, Rubio, Hontangas, & Hernández, 2005; Rubio, Aguado, Hontangas, & Hernández, 2007). This computer-based questionnaire, designed as an adaptive test, consists of 28 items. All items have a graded response option from 1 (*totally agree*) to 4 (*totally disagree*). An example of an EAB item is *My muscles are usually tense*. Due to the fact that π* works under contingency tables with the same number of cells as response options raised to the power of the number of items, the item score was dichotomized and only 8 items were used. The first one was eliminated to avoid any individuals' adjustment effect to the test and items 2 to 9 were selected. Moreover, options 1 and 2 (*Totally agree* and *Agree*) were joined as were options 3 and 4 (*Disagree* and *Totally disagree*). Thus, the number of possible response patterns was 2^8, equal to 256 cells. No test had an empty cell. Working with a larger number of items or response options would increase the probability of empty cells which would make the application of the model impossible (for a detailed description of the model, see Hernández et al, 2006).

EAB used as a linear test (as in the present study) shows a Cronbach's α = .91, an adequate convergent validity using the BFQ EA scale (Caprara, Barbanelli, & Borgogni, 1993) and the EPQ N scale (Eysenck & Eysenck, 1975) (r = .77 and r = .86, respectively; see Aguado et al, 2005, Rubio et al., 2007). The 8-item reduced version of EAB showed a Cronbach's α = .76, whilst the Cronbach's alpha for the rest of the items was α = .89. Correlations between the 8-item version, the full version and the version with the rest of items were r = .68 (r corrected for unreliability = .92), and r = .63 (r corrected for unreliability = .93), respectively.

Thoroughness Test (TT) (Hernández, Sánchez-Balmisa, Madrid, & Santacreu, 2003). This task-based test assesses the thoroughness in identifying and marking a series of objects equal to a target (specifically, a type of tree) in a screen matrix containing those objects among others (other types of trees). Each item presents a matrix of 12 rows by 10 columns (120 cells) and 60% of the cells (72) contain a figure. Of these 72, 14 contain the target element whereas 58 have a distractor (different types of trees). The elements to identify are distributed in 7 rows and 7 columns, with two target elements each. The test has 15 trials. Each trial lasts 20 seconds. Afterwards, a new matrix appears. The score is based on the order in which objects are identified. The participant gains a point each time he/she consecutively marks two elements in the same row/column; otherwise he/she collects 0. Thus the score ranges from 0 (lower degree of thoroughness) to 7 (higher thoroughness), multiplied by the 15 trials (see Hernández, et al., 2003). For the same reasons as above, the item score was dichotomized, grouping values 0 to 3 in the first category and values from 4 to 7 in the second. Using the same criterion as above, only 2-9 items were used. The TT shows an internal consistency of Cronbach's α = .94 and convergence validity of r = .64 using another objective task-based test assessing Thoroughness (Hernández, Lozano, Shih, & Santacreu, 2009). The reduced version showed a Cronbach's α = .69, and Cronbach's alpha for the rest of items was α = .67. Correlations between the 8-item version, the full version and the version with the rest of items were r = .81 (r corrected by unreliability = .97), and r = .70 (r corrected for unreliability = .85), respectively

Procedure

The assessment tests were included as a part of those used in the selection process. The total length of the assessment process was three hours and it was entirely proctored on the premises of the agency in charge. No previous computer skills were required (only the use of the mouse). Performance of each participant was automatically registered on a database for further calculation.

Results

Table 1 shows the estimated values of the 8-item a_i (discrimination) and b_i (item location) parameters, the goodness of fit of the model (G^2), degrees of freedom (df), probability (p), and the statistic (π*) which represents the estimated proportion of inconsistent individuals, as well as the lower bound of π* (π_L*) for both tests. G^2/df ratio for

Table 1
Results for Dichotomous Data. Item parameter estimates and goodness of fit statistics for the 2PL model

	EAB		TT	
Item	*a*	*b*	*a*	*b*
1	1.970	- .402	.555	-2.770
2	1.934	- .707	.518	4.382
3	1.839	.217	.898	- .015
4	1.911	-1.283	.410	2.462
5	2.152	- .362	1.611	-1.057
6	1.520	- .885	.700	1.036
7	2.347	-1.173	.929	1.597
8	1.824	- .841	.356	.822
G^2	359.10		323.14	
df	230		239	
p	.000		.000	
π*	.412		.504	
π_L*	.347		.375	

Note. EAB is the Emotional Stability questionnaire; TT is the task-based test for assessing Thoroughness; *a* and *b* are the model parameters; G^2 is the goodness of fit statistic, *df* are the degrees of freedom, *p* is the p-value, π* is the proportion of subjects out of the model and π_L* is the lower bound of π*.

https://doi.org/10.5209/rev_SJOP.2011.v14.n1.18 Published online by Cambridge University Press

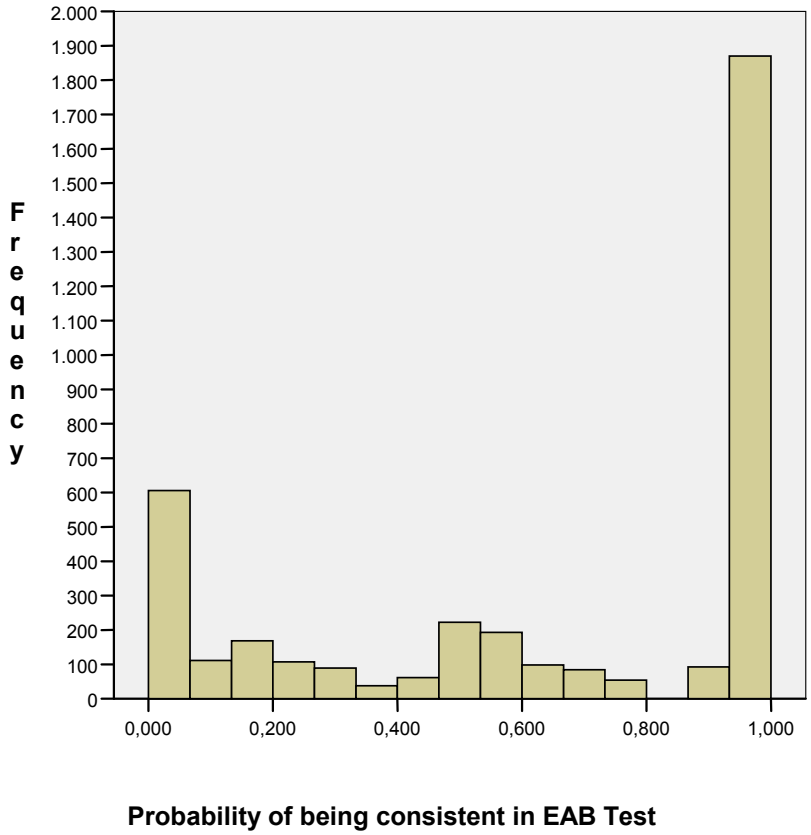

Figure 1. Distribution of the variable "probability of being consistent" according to the emotional stability questionnaire EAB (from 0 to 1).

EAB and for TT are in the 1-3 range which indicates an adequate goodness of fit of the model, (see, for instance, Carmines & McIver, 1981).

According to the π* statistic, the estimated percentage of consistent individuals in the EAB was 59%, whereas for Thoroughness it was 49.6%. This result points out that consistency shown in the self-report is higher than in the task-based test.

From direct tests scores, the probability of each participant being consistent (*p-consistent*) was computed for each test using the procedure described in Hernández et al. (2006). Thus, a new variable was generated. This variable ranged from 0 (minimum probability of consistency) to 1 (maximum probability of consistency). In order to test whether there was a systematic relationship between trait level and the probability of being consistent, the correlations between *p-consistent* and θ was computed. The results were $r = .42$ for EAB and $r = -.19$ for TT, respectively. Moreover, a quadratic regression analysis was carried out. The results were $R^2 = .38$ for EAB and $R^2 = .07$ for TT, respectively, slightly higher than linear relationships. These results are in keeping with Paunonen and Jackson (1985)'s who pointed out that a high degree of behavioral variability (low level of consistency) for a trait would preclude at an extreme level of a trait. Nevertheless, they show there was no bias due to the latent trait level. The distributions of this variable in the EAB and the TT are shown in Figures 1 and 2 respectively.

As can be seen, both are non-normally distributed, presenting a type of trimodal distribution. A tentative interpretation of these results would be that the distribution is composed of three different groups. On the one hand, there is an important group of participants who appear highly consistent; on the other hand, there is another important group, though smaller than the first, of highly inconsistent participants. Thirdly, there is a group which ranges from one pole to the other.

Correlation between *p-consistent* for each test was, ($r = .01$, $p = .92$). Moreover, the Wilcoxon test shows that the null hypothesis should be rejected ($Z = -3.66$, $p < .0001$). According to these results there is no relationship between the distributions.

Following this, participants were categorized according to the central point of the theoretical distribution. Individuals with values higher than .50 in probability of being consistent were included in the consistency group; those with values lower or equal to .50 were included in the inconsistent group. Table 2 shows the percentage of individuals classified according to this criterion.

https://doi.org/10.5209/rev_SJOP.2011.v14.n1.18 Published online by Cambridge University Press

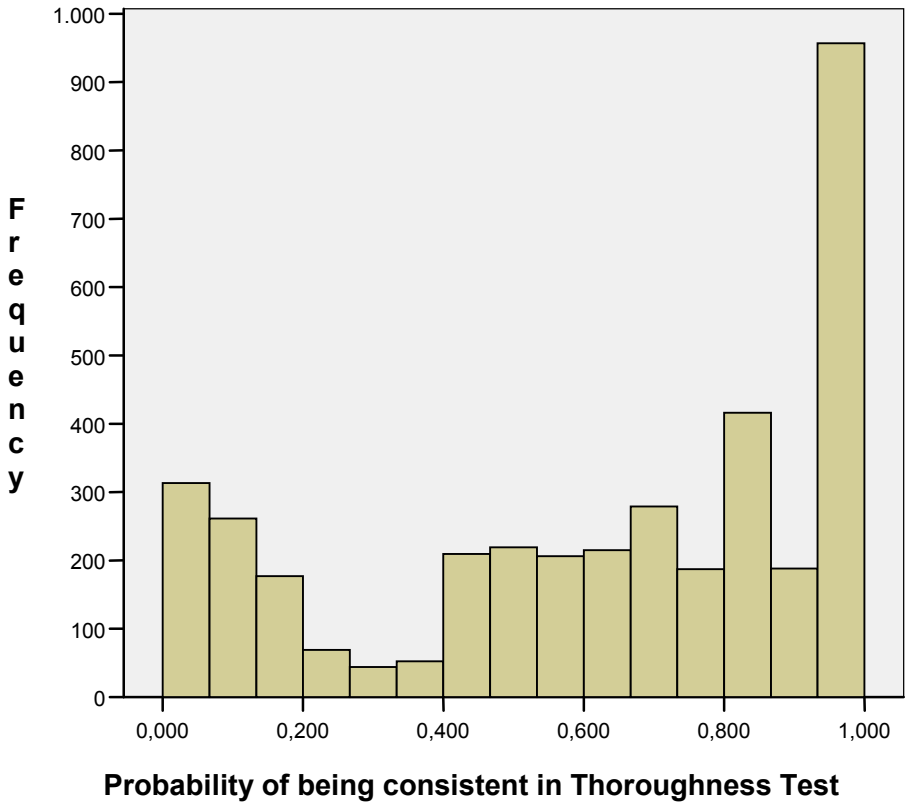

Figure 2. Distribution of the variable "probability of being consistent" according to the task-based Conscientiousness Test (from 0 to 1).

Table 2
Percentage of consistent (individuals with probability of being consistent p-consistent > .50) and inconsistent (individuals with probability of being consistent p-consistent ≤ .50) participants in both EAB and TT

	Inconsistent	Consistent	Total
Inconsistent.	425 *(11.2%)*	795 *(21.0%)*	1,220 *(32.2%)*
Consistent	881 *(23.2%)*	1,691 *(44.6%)*	2,572 *(67.8%)*
Total	1,306 *(34.4%)*	2,486 *(65.6%)*	3,792 *(100%)*

Note. EAB is the Emotional Stability questionnaire; TT is the objective task-based Thoroughness test.

Taking into account that the criterion used was less restrictive, marginal values of non consistency are closer to the lower bound of the interval (π_L*) observed in Table 1. Thus, it can be seen that using this criterion, the percentage of consistent individuals is 67.8% for the questionnaire and 65.6% for the task-based test. A t-test comparing the mean between both distributions showed significant differences ($Mean_{EAB}$ = .65; Sd_{EAB} = .40; $Mean_{TT}$ = .62; Sd_{TT} = .33; t = 2.96; df = 3,791; p = .003) Moreover, the McNemar test for correlated proportions showed the proportion of consistent individuals is greater in EAB than in TT (χ^2 = 431; df = 1; p = .038). The contingency table shows four differentiated groups. 44.6% of participants were consistent in both tests (they have a value higher than

https://doi.org/10.5209/rev_SJOP.2011.v14.n1.18 Published online by Cambridge University Press

.50 in both tests). 11.2% were inconsistent in both tests. On the other hand, a similar percentage of consistencies were revealed in one but not in the other test: 21.0% were consistent in just the task-based tests and 23.2% were consistent only in the self-report test, as emerged from the very low correlation coefficient.

Discussion

Firstly, the results demonstrate a percentage of individuals who fall outside the model, both when using the self-report instrument (Q-data in Cattell's terminology) or an objective task-based test (T-data). In other words, according to the procedure for determining individual consistency based on the invariance of θ throughout the test, there are a remarkable number of individuals who would not be considered consistent (30-35%). This finding agrees with previous results obtained by Hernández et al. (2006).

Secondly, the results obtained support the hypothesis that people are differentially consistent depending on whether assessed by a questionnaire or by an objective task-based test. In other words, individual consistency differs according to whether individuals are qualifying themselves using verbal statements or whether their stylistic way of carrying out a task is assessed. Moreover, individuals who are consistent in the self-report are not always the same as those who are consistent in the task-based test. These results support the idea that the way individuals synthesize their experience would be different when dealing with verbal propositions and when behaving (Santacreu et al., 2006). The finding that people are more consistent when talking about themselves relates to human beings developing discourse patterns which are essentially coherent (Mischel, 2004; Shoda & Mischel, 2000). Individuals try not to contradict themselves and attempt to preserve a coherent meaning, organizing their experiences into meaningful life stories (Cervone & Shoda, 1999; Greenwald, 1980; Ross, 1989). Furthermore, this coherence must play a fundamental role for preserving the psychological integrity of the person (see McLean, et al., 2007, for the role of life stories in the construction of the self).

Thirdly, results revealed that there are individuals who show consistency in both variables, those who show consistency in just one of them, and, moreover, there are an important number of individuals who show consistency in neither one nor the other. Therefore, individuals do not generally behave consistently or inconsistently. An individual's consistency depends on personality dimension or on the type of assessment instrument (self-report or objective task-based test), the number of consistent individuals in the self-report test being higher than in the objective task-based test.

From our point of view, there are significant consequences of these findings: 1) *Regarding the predictability of future behaviour*. The number of inconsistent individuals directly affects the predictive capacity of personality assessment instruments. Obviously, it will be more accurate to predict future behaviours of those consistent as opposed to those that are inconsistent. Moreover, determining individual consistency will improve the predictive validity of the instrument. At this point, it should be noted that the tests used in the present study do not assess the same dimension. Thus, it is not possible to compare individual's consistency in the same personality dimension. Further studies would allow the comparison of individual's consistency in self-reports and task-based tests within the same dimension in order to provide data on individuals' consistency in a personality dimension irrespectively of the type of test. 2) *Concerning the psychometric principles of the assessment instruments*. It is expected that a personality measure shows a high degree of statistical consistency, both in terms of internal consistency (item homogeneity) and in terms of reliability (degree of consistency in participant scores in test replications). Those items which do not contribute to the internal consistency of the test are removed as they do not measure the dimension in question; those tests which show significant differences between replications are ruled out as personality measures. Nevertheless, there are at least two sources of variation in a test score: items and individuals assessed. If the time between applications, assessment conditions, examiner, etc. is controlled for, differences between these two replications can usually be attributed to the assessment instrument, however, this does not change either. To reject the origin of the variability being attributable as much to the tests as to the individual (or any combination of these) is illogical reasoning. In fact, items are the same for everyone and they are the same in the previous administration and in the replication. Excluding voluntary biases, differences might be attributed to: a) a change in how the person considers him/herself or behaves; b) moderator variables, such as the way individuals interpret the items/situations, play a role; c) no correlation between the previous response and the personality dimension assessed; and, d) individuals present a lack of consolidation of the personality dimension which determines the lack of consistency showed in the test. The result obtained reaffirms that the values of individual consistency should be taken into account when establishing psychometric properties of an assessment instrument. Methods based on IRT models such as the π^* allow to separate these two sources of variation.

Even though it was not the main objective of the paper, the procedure used has demonstrated its utility for classifying individuals. Thus, it would provide a useful alternative way for determining traitedness (Baumeister & Tice, 1988; Bem & Allen, 1974; Cucina & Vasilopoulos, 2005). Future studies should compare this procedure with the traditional measures based on the ipsatized variance indexes usually used for these purposes.

https://doi.org/10.5209/rev_SJOP.2011.v14.n1.18 Published online by Cambridge University Press

References

Aguado, D., Rubio, V. J., Hontangas, P., & Hernández, J. M. (2005). Propiedades psicométricas de un test adaptativo informatizado para la medición del ajuste emocional [Psychometric properties of a computer-adaptive-test for assessing emocional adjustment]. *Psicothema*, *17*, 484-491. doi:10.1027/1015-5759.23.1.39

Barrick, M. R., & Mount, M. K. (1991). The Big-Five personality dimensions: A meta-analysis. *Personnel Psychology, 44*, 1-26. doi:10.1111/j.1744-6570.1991.tb00688.x

Baumeister, R. F., & Tice, D. M. (1988). Metatraits. *Journal of Personality, 56*, 571-598. doi:10.1111/j.1467-6494.1988.tb00903.x

Bem, D. J., & Allen, A. (1974). Predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review, 81*, 506-520. doi:10.1037/h0037130

Biesanz, J. C., & West, S. G. (2000). Personality coherence: Moderating self-other profile agreement and profile consensus. *Journal of Personality and Social Psychology, 79*, 425-437. doi:10.1037//0022-3514.79.3.425

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick, *Statistical theories of mental test scores.* (pp. 397-472) Reading, MA: Addison-Wesley.

Bromiley, P., & Curley, S. P. (1992). Individual differences in risk taking. In J. F. Yates, (Ed.). *Risk-taking behavior* (pp. 87-132). New York, NY: John Wiley.

Caprara, G. V., Barbaranelli, C., & Borgogni, L. (1993). *BFQ: Big five questionnaire. manuale.* Firenze: Organizzazioni Speciali.

Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables. Anaylisis of covariaces structures. In G. W. Bohrnstedt, & E. G. Borgatta (Eds.). *Social measurement. current issues* (pp. 65-115). Beverly Hills, CA: Sage. doi:10.1207/s15327965pli0404_1

Caspi, A., & Moffitt, T. E. (1993). When do individual differences matter? A paradoxical theory of personality coherence. *Psychological Inquiry*, *4*, 247-271. doi:10.1207/s15327965pli0404_1

Cattell, R. B. (1965). *The scientific analysis of personality*. Baltimore, MD: Penguin.

Cattel, R. B., & Kline, P. (1977). *The scientific analysis of personality and motivation*. London: Academic Press.

Cattell, R. B., & Warburton, E. W. (1967). *Objetive personality and motivation tests*. Urbana, IL: University of Illinois Press.

Cervone, D., & Shoda, Y. (1999). Social-Cognitive Theories and the Coherence of Personality. In D. Cervone, & Y. Shoda (Eds.). *The coherence of personality: Social-cognitive bases of consistency, variability and organization* (pp. 3-33). New York, NY: Guilford.

Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality, 59*, 143-178. doi:10.1111/j.1467-6494.1991.tb00772.x

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and serf-opinion. *Personality and Individual Differences, 5*, 11-26. doi:10.1016/0191-8869(84)90133-8

Cucina, J. M., & Vasilopoulus, N. L. (2005). Nonlinear personality-performance relationships and the spurious moderating effects of traitedness. *Journal of Personality, 73*, 227-259. doi:10.1111/j.1467-6494.2004.00309.x

Drasgow, F., Levine, M., & Williams, E. A. (1985). Appropriateness measurement with politomous item response models and standardized indexes. *British Journal of Mathematical and Statistical Psychology, 38*, 67-96.

Epstein, S. (1985). The person-situation debate in historical and current perspective. In E. E. Roskam (Ed.). *Measurement and personality assessment* (pp. 315-321). Amsterdam: North Holland.

Eysenck, H. J, & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. San Diego, CA: Educational and Industrial Testing Service.

Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual Differences*, *34*, 1225-1243. doi:10.1016/S0191-8869(02)00111-3

Ferrando, P. J. (2004). Person reliability in personality measurement. An item response theory analysis. *Applied Psychological Measurement, 28*, 126-140. doi:10.1177/0146621603260917

Funder, D. C. (1983). The consistency controversy and the accuracy of personality judgement. *Journal of Personality, 5*, 346-359. doi:10.1111/j.1467-6494.1983.tb00337.x

Greenwald, A. G (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*, 603-618. doi:10.1037/0003-066X.35.7.603

Hellervik, L. W., Hazucha, J. F., & Schneider, R. J. (1992). Behavior change: Models, methods, and a review of the evidence. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed), (Vol. 3, pp. 821-895). Palo Alto, CA: Consulting Psychologists Press.

Hernández, J. M., Lozano, J. H., Shih, P. C., & Santacreu, J. (2009). Validez convergente de dos pruebas objetivas de personalidad que evalúan minuciosidad [Convergent validity between objective assessment tests of interactive style Thoroughness]. *Psicothema*, *21*, 133-140.

Hernández, J. M., Rubio, V. J., Revuelta, J., & Santacreu, J. (2006). A procedure for estimating intra-subject consistency of behavior. *Educational and Psychological Measurement, 66*, 417-434. doi:10.1177/0013164405275667

Hernández, J. M., Sánchez-Balmisa, C., Madrid, B., & Santacreu, J. (2003). La evaluación objetiva de la minuciosidad: Diseño de una prueba conductual [The objective assessment of thoroughness: The design of a behavioral test]. *Análisis y Modificación de Conducta*, *29*, 457-479.

Hofmann, W., & Schmitt, M. (2008). Advances and challenges in the indirect measurement of individual differences at age

10 of the Implicit Association Test. *European Journal of Psychological Assessment, 24*, 207-209. doi:10.1027/1015-5759.24.4.207

Lang, P. J. (1971). The application of psychophysiological methods to the study of psychotherapy and behaviour modification. In A. E. Bergin, & S. L. Garfield (Eds.). *Handbook of psychotherapy and behavior change* (pp. 75-125). New York, NY: Wiley.

Lanning, K. (1988). Individual differences in scalability: An alternative conception of consistency for personality theory and measurement. *Journal of Personality and Social Psychology, 55*, 142-148. doi:10.1037/0022-3514.55.1.142

Linehan, M. M., & Kehrer, C. A. (1993). Borderline personality disorder. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders* (2nd ed.) (pp. 396-441). New York, NY: Guilford Press.

McCrae, R. R., & Costa, P. T. (1994). The stability of personality: Observation and evaluations. *Current Directions in Psychological Science, 3,* 173-175. doi:10.1111/1467-8721.ep10770693

McLean, K. C., Pasupathi, M., & Pals, J. L. (2007). Selves creating stories creating selves: A process model of self-development. *Personality and Social Psychology Review, 11*, 262-278. doi:10.1177/1088868307301034

Mischel, W. (1968) *Personality and assessment*. New York, NY: Wiley.

Mischel, W. (2004). Toward and integrative science of the person. *Annual Review of Psychology, 55*, 1-22. doi:10.1146/annurev.psych.55.042902.130709

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, *49*, 229-258. doi:10.1146/annurev.psych.49.1.229

Ozer, D. (1986). *Consistency in personality*. Berlin: Springer-Verlag.

Paunonen, S. V., & Ashton, M. C. (2001a). Big Five factors and facets and the prediction of behaviour. *Journal of Personality and Social Psychology*, *81*, 524-539. doi:10.1037//0022-3514.81.3.524

Paunonen, S. V., & Ashton, M. C. (2001b). Big Five predictors of academic achievement. *Journal of Research in Personality*, *35*, 78-90. doi:10.1006/jrpe.2000.2309

Paunonen, S. V., Haddock, G., Forsterling, F., & Keinonen, M. (2003). Broad versus narrow personality measures and the prediction of behavior across cultures. *European Journal of Personality*, *17*, 413-433. doi:10.1002/per.496

Paunonen, S. V., & Jackson, D. N. (1985). Idiographic measurement strategies for personality and predictions: Some unredeemed promissory notes. *Psychological Review, 92*, 486-511. doi:10.1037/0033-295X.92.4.486

Pawlik, K. (1985). Cross-situational consistency behavior: models, theories, and in-field tests of the consistency issue. In E. E. Roskam (Ed.). *Measurement amd personality assessment*. (pp. 307-314) Amsterdam: North Holland.

Pervin, L. A. (1996). *The science of personality*. New York, NY: Wiley.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3-25. doi:10.1037//0033-2909.126.I.3

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341-357. doi:10.1037/0033-295X.96.2.341

Rubio, V. J., Aguado, D., Hontangas, P., & Hernández, J. M. (2007). Psychometric properties of an emotional adjustment measure. *European Journal of Psychological Assessment*, *23*, 39-46. doi:10.1027/1015-5759.23.1.39

Rudas, T., Clogg, C. C., & Lindsay, B. G. (1994). A new index of fit based on mixture methods for the analysis of contingency tables. *Journal of the Royal Statistical Society, series B*, *56*, 623-639.

Rudolph, A., Schröeder-Abe, M., Schütz, A., Gregg, A. P., & Sedikides, C. (2008). Through a glass, less darkly? Reassessing convergent and divergent validity in measures of implicit self-esteem. *European Journal of Psychological Assessment, 24*, 273-281. doi:10.1027/1015-5759.24.4.273

Rusthon, J. P., Jackson, D. N., & Paunonen, S. V. (1981). Personality: Nomothetic or idiographic? A response to Kenrick & Stringfield. *Psychological Review, 88*, 582-589. doi:10.1037/0033-295X.88.6.582

Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology. 82*, 30-43. doi:10.1037/0021-9010.82.1.30

Santacreu, J., Rubio, V. J., & Hernández, J. M. (2006). The objective assessment of personality: Cattell's T-data revisited and more. *Psychology Science*, *48*, 53-68.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274. doi:10.1037/0033-2909.124.2.262

Semin, G. R., & Krahe, B. (1988). Explaining perceived cross-situational consistency: Intuitive psychometrics or semantic mediation. *European Journal of Personality*, *2*, 239-252. doi:10.1002/per.2410020402

Shoda, Y., & Mischel, W. (2000). Reconciling contextualism with the core assumptions of personality psychology. *European Journal of Personality, 14*, 407-428. doi:10.1002/1099-0984(200009/10)14:5<407::AID-PER391>3.0.CO;2-3

Skinner, N. S. T., & Howarth, E. (1973). Cross-media independence of questionnaire and objective test personality factors. *Multivariate Behavioral Research, 8*, 23-40. doi:10.1207/s15327906mbr0801_2

Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108-131. doi:10.1207/S15327957PSPR0402_01

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247. doi:10.1207/s15327957pspr0803_1

Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1994). Meta-analysis of Personality-Job performance relatons: A reply to Ones, Mount, Barrick and Hunter (1994). *Personnel Psychology*, *47,* 1157-1172. doi:10.1111/j.1744-6570.1994.tb02415.x

Wagerman, A. S., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality 41*, 221-229. doi:10.1016/j.jrp.2006.03.001

Weiner, B. (1990). Attributions in applied personality theory. In L. A. Pervin (Ed.). *Handbook of personality psychology* (pp. 465-485). New York, NY: Guilford Press.

West, S. G., & Graziano, W. G. (1989). Long-term stability and change in personality: An introduction. *Journal of Personality, 57,* 175-193. doi:10.1111/j.1467-6494.1989.tb00479.x

Zuckerman, M. (1979). Traits, status, situations, and uncertainty. *Journal of Psychopathology and Assessment*, *1*, 43-54. doi:10.1007/BF01322417

Received January 29, 2009
Revision received January 19, 2010
Accepted January 25, 2010